



Assessing contemporary legislative proposals for their compatibility with a natural law case for AI legal personhood

Joshua Jowitt¹

Received: 24 December 2019 / Accepted: 9 April 2020
© The Author(s) 2020

Abstract

The question of the moral status of AI and the extent to which that status ought to be recognised by societal institutions is one that has not yet received a satisfactory answer from lawyers. This paper seeks to provide a solution to the problem by defending a moral foundation for the recognition of legal personhood for AI, requiring the status to be granted should a threshold criterion be reached. The threshold proposed will be bare, noumenal agency in the Kantian sense. Agency has been identified by Alan Gewirth as the source of the rights claims of our own species and, at risk of contradiction, is a foundation that must be expanded to all agents or else we contradict the foundation of our own rights. This is something that ought to be recognised through the granting of legal personhood to all noumenal agents by any system that requires such personhood for the enforcement of rights, or else the rule restricting legal personhood cannot be seen as a valid legal norm. Having laid out the case, the paper will move on to defend this natural law conception against the narrower definition of legal personhood proposed by Bryson et al. with regards to AI. It will argue that bare agency is a sufficient, though not necessary, criterion for the ascription of legal personhood in any system that sees the status as necessary for the ascription of legal rights. The paper will conclude by analysing the proposals currently making their way through the legislatures of the UK and European Union. They will be assessed for their compatibility with the claim that a functioning legal system necessarily must recognise the legal personhood of all noumenal agents regardless of their origins, and whether they are future-proofed for the possibility that AI may meet this threshold.

Keywords Legal personhood · Natural law · AI · Rights · Gewirth

1 Introduction

The question of whether AI should be granted legal personhood is one that has received prior scholarly attention, but has not yet been satisfactorily answered. Authors from Smith to Solum and beyond have attempted to define what it means to be a legal person and apply their findings to the issues raised by AI,¹ but a problem faced by those interested in the question is that the concept of legal personhood is itself a contested idea, with comparatively little scholarship on its substantive boundaries or theoretical foundation beyond the orthodox viewpoint that the status acknowledges the bearer as ‘[t]he subject of rights and duties.’² As a result, the possibility of granting of legal personhood to AI is a

dilemma whose resolution is more elusive than might first be expected.

The purpose of this paper is to defend a metaphysical conception of legal personhood grounded in the moral philosophy of Alan Gewirth. This conception has at its foundation the agency of the subject, and it will be proposed that any AI that possesses the noumenal agency required to be a moral patient for the purposes of Gewirth’s theories must be granted legal personhood by any legal system that sees legal personhood as necessary for the enforcement of legal rights. It will be argued that failure to do so would result in a legal rule that is incapable of providing the absolute reason for action that any law must necessarily claim to be capable of making. In providing a foundation for legal personhood grounded in a strong theory of Natural Law this paper will demonstrate that legal concepts and their normative force are inescapably grounded in moral norms. It is impossible

✉ Joshua Jowitt
joshua.jowitt@newcastle.ac.uk

¹ Newcastle Law School, 21-24 Windsor Terrace,
Newcastle upon Tyne NE1 7RU, UK

¹ Smith (1928), Solum (1992).

² Smith, ‘Legal Personality’ (n 1) 283.

for legal personhood to be seen as a thing separate from a metaphysical criterion of personhood, and claims that argue otherwise are demonstrably false.

Having established the natural law case for legal personhood for AI that meet the threshold criterion of agency, the paper will move on to critique a contemporary denial of legal personhood for AI found in the writing of Bryson et al., which will be seen to be flawed for three reasons, each connected to their reluctance to view metaphysical conceptions of personhood unavoidably part of the legal domain. The paper will conclude by examining current legislative proposals before the UK and EU legislatures for their compatibility with the Gewirthian legal personhood proposed in earlier sections, asking whether they succeed in future proofing themselves against the possibility that AI may reach the threshold criterion required.

2 The principle of generic consistency (PGC)

This paper adopts a strong natural law position that defends the view that legal normativity is inescapably grounded in morality. This claim requires the existence of an objective, rationally identifiable moral standard against which the validity of legal norms can be assessed. In the PGC, Alan Gewirth claims to have identified such a supreme moral principle; it behaves as a categorical imperative against which the moral permissibility of action can be identified³ and, by extension, the validity of legal norms and constructs can be assessed. It is not within the scope of this paper to mount a defence of the PGC.⁴ Rather, an outline will be offered that adopts Beyleveld's subdivision of the argument into three steps for the clarity that it brings to the dialectically necessary progression of the PGC⁵ before applying the argument to the concept of legal personhood.

The first stage begins with a statement of what it means to be a prospective purposive agent (hereafter PPA)—that I perform an action of my own volition to bring about a given end. This establishes the standard philosophical understanding of agency as the starting point of the PGC, and was chosen by Gewirth because of this alignment.⁶ Stage one entails several implicit and interconnected statements—each of

which, at this point, remain value neutral. Firstly, the agent necessarily sees the end being pursued as being subjectively desirable or otherwise of benefit to them. If this were not true, the agent would not have a reason to attempt to achieve the end in question. Secondly, the agent necessarily requires a minimum level of both freedom and wellbeing to achieve the end. This is axiomatic of any purposive voluntary action beyond natural impulse of reflex, as such an action cannot take place if the agent does not enjoy a bare minimum level of personal autonomy and physical and mental wellbeing. Finally, the instrumental necessity of these conditions to all action requires the agent to recognise them as generic conditions of agency (hereafter GCAs).⁷

Stage two builds from the identification of these GCAs. If GCAs are necessary for any purposive action, and purposive action is necessary to attain any end that might be achieved, then an agent must—from their own internal point of view—see non-consensual interference with their GCAs as undesirable. Thus, the agent is committed to seeing themselves as possessing a claim-right that their GCAs ought not to be interfered with.⁸ This conclusion is a logical consequence of stage one, thus making the dialectical necessity of the argument apparent; if an agent must recognise that their GCAs are instrumentally necessary goods that are universally necessary for any purposive action, itself necessary for any end to be successfully attained, then that they must make claim-rights to their GCAs is a logical consequence of the standard view of agency itself.

Stage three proceeds from this conclusion to universalise the rights-claim made from the internal viewpoint of an agent. If I am committed to making a claim-right to my GCAs because of the necessary and unavoidable starting point of my agency, then all other PPAs must be capable of making the same rights claim. Should I refuse to acknowledge that other PPAs are capable of making the same rights claim as myself, I undermine my own rights-claim by suggesting that another PPA cannot claim rights as a result of their agency despite agency being the source of my own rights-claim.⁹ This should be rejected as logically inconsistent, as such a rejection entails the denial of one of two truths established through the dialectical necessity of the argument's progression: denial either that I myself have rights to my GCAs by removing the necessary grounding of rights in agency; or of my own agential status. As both of these denials require the me to use my own agency and thus use my GCAs, a paradox is created where the criterion whose necessity they seek to reject is itself necessary for the denial to be

³ Gewirth (1978, 1982, 1996); Gewirth, 'Replies to my Critics' in Regis (1984).

⁴ This has been comprehensively undertaken elsewhere see Beyleveld (1991).

⁵ Beyleveld, *The Dialectical Necessity of Morality* (n 4) 14.

⁶ Engaging with the debate as to whether this is the correct view of what it means to be an agent is not within the purview of this paper. It instead assumes that, in the absence of convincing evidence to the contrary, it should be accepted as an accurate representation of what it means to be an agent.

⁷ Gewirth, *Reason and morality* (n 3) 22–63.

⁸ Gewirth 63–103.

⁹ Gewirth 104–198, esp. 104–128; Beyleveld, *The dialectical necessity of morality* (n 4) 44–45.

successful. The argument against universalisation therefore fails, and all PPAs are required to respect the claim rights of all other PPAs. A normative obligation has therefore been created whereby permissible behaviour amongst agents is restricted to that which does not non-consensually interfere with other agents' claim rights to their GCAs.

Should AI possess such noumenal agency, it would be a moral patient for the purposes of the PGC. Yet ascertaining whether this is the case is problematic for the observer. Knowing with certainty that a being other than ourselves is a noumenal agent is impossible; at best, we can observe and decide whether a being behaves as though they are an agent.¹⁰ But faced with the maxim behaving as a categorical imperative that results from the PGC, such fence-sitting is undesirable. As we are required to avoid breaching its requirements whenever possible, a precautionary principle comes into play.¹¹ Such a principle exists where we observe a being whose agential status is uncertain, yet who behaves in a way that suggests they may possess noumenal agency; a being that appears to act with purposiveness and voluntariness manifest through seeming to undertake action X for purpose E.¹² In such a situation, the following statement is true:

- (a) If the being is an agent, then:
 - (i) If the observer treats them in a way which is PGC compliant, then the observer's behaviour is morally permissible.
 - (ii) If the observer treats them in a way which is not PGC compliant, then the observer's behaviour is morally impermissible.
- (b) If the being is not an agent, then:
 - (i) If the observer treats them in a way which is PGC compliant, then the observer's behaviour is morally permissible.
 - (ii) If the observer treats them in a way which is not PGC compliant, then the observer's behaviour is morally permissible.

Given the normative requirement that we must avoid behaviour that is not PGC compliant, the only course of action that must categorically be avoided is (a-ii). The only

way to be certain that we avoid this outcome when the agential status of the being in question is unknown is to always treat the being in a way which is PGC compliant. Therefore, if the agential status of a being in question is unknown but we have reason to believe that they are behaving as though they are an agent, we should err on the side of caution and presume that they are. This is the only way to be certain that we avoid breaching the requirements of the PGC. The question before us is therefore not whether or not the AI before us is an agent, but whether or not it possesses sufficient evidence of agency so as to lead us to believe that it could be an agent. If so, it has valid moral claim rights as per the PGC.

Having reached this conclusion, we can turn to the question of legal personhood for AI. The argument to follow holds that the claim-rights arising from the moral patienthood of AI that meet the threshold criterion should be seen as sufficient for the recognition of legal personhood, should that status be necessary in a legal system for legal rights to be recognised, adjudicated upon and enforced.

3 The PGC and legal personhood

To summarise briefly, the argument for the PGC presented above applies to all PPAs that, subject to the precautionary principle, we have reason to believe are capable of acting purposefully and voluntarily beyond reflex or natural impulse, and concludes with the dialectically necessary observation that the rights claims that result from the PGC are of moral concern. Given it begins from the internal viewpoint of the agent, it differs from other tests for moral patienthood in that it is based on a criterion that we possess ourselves¹³ rather than an attempt to locate a characteristic of moral worth in beings to which we may wish to ascribe rights. A failure to ascribe rights to agents that are capable of making the same rights claims as those that we necessarily claim for ourselves entails the logical impossibility of denying the importance of our own agency whilst using said agency to undertake the denial; this is a contradiction as we must necessarily value our agency and, by extension, our GCAs, when undertaking this—and all—action. Gewirth has therefore identified a principle that behaves in the same way as a categorical imperative that requires all agents to respect the GCAs of all other agents.

But should this moral concern extend to legal personhood for PPAs that are AI? This paper seeks to argue that legal personhood for such PPAs is not optional, but necessary in any internally coherent legal system. To demonstrate this admittedly bold claim, a necessary connection between the PGC and legal normativity must be identified against

¹⁰ Beyleveld (2012), 9–10; for an account of the general problem of the unknowability of other minds see Nagel (1974), 436–440.

¹¹ The precautionary principle is well recognised amongst various ethical traditions. For a full defence of the desirability of acting according to a precautionary principle when the moral status of a being is uncertain, see Sebo (2018).

¹² Gewirth, Reason and morality (n 3) 27–37, 43.

¹³ Beyleveld, The dialectical necessity of morality (n 4) 44–45.

which the question of the necessity of such recognition can be posed. Any such enquiry as to the nature of legal rules requires us to identify a definition of law as starting point to our enquiry that is equally acceptable to both positivists and natural lawyers requires us to look at the function of law for our starting point.¹⁴ Beyleveld and Brownsword have argued persuasively that such a starting point can be found by beginning with the function and purpose of law—the subjection of human conduct to the governance of rules. This functional account is a necessary starting point as to make a claim as to the source of the norm in question at this stage would presuppose either a positivist/natural law starting point and, by extension, presuppose a position in the dispute that it seeks to resolve.¹⁵

Regardless of its source, it is axiomatic that a rule claiming to be law must claim to be able to direct those to whom it is addressed, and equally necessary that the rule must claim to be absolute and override all other competing reasons for action. These conjoined claims are necessary as, should a law not claim to be an absolute reason capable of guiding action, it would be unable to guide the action of those who were not already predisposed to follow it. The necessity of this claim to be an absolute reason locates legal norms within a unified conception of practical reason; were this incorrect, the rule claiming to be law would be incapable of claiming to override all other conflicting reasons. And in thus situating itself within a unified conception of practical reason, law can be seen to operate on the same normative plane as moral reasons for action.¹⁶

This location requires us to establish how this claim to be an overriding reason for action interacts with the PGC. The PGC acts in the same way as a categorical imperative to all agents, and therefore is definitionally located at the top of any normative hierarchy of reasons for action. The consequence of this is that any direction which is not PGC compliant is *prima facie* incapable of providing an absolute reason to be followed. By extension, any rule pertaining to be law that is not PGC compliant is incapable of claiming to be an absolute reason to guide the action of those to whom it is addressed as any normative claim made by the rule would be superseded by the contradictory requirements of the PGC. Such a rule would therefore definitionally fail to be law, as

it would be incapable of subjecting human conduct to its requirements.¹⁷

Having established that legal rules must be PGC compliant or else they are incapable of providing an absolute reason to guide human conduct, we can attempt to connect this conclusion to the concept of legal personhood. Legal personhood can trace its origins to Roman law and is commonly accepted as being a construct designed denote that the legal person is the beneficiary of a bundle of legally enforceable rights.¹⁸ Despite recent scholars casting doubt on the extent to which this orthodoxy remains, if it ever was, a valid understanding of the concept,¹⁹ the position that legal personhood is necessary for the recognition of legal rights remains widely held. Previous work has been undertaken that explores the connection between the PGC and the concept of legal personhood,²⁰ but space precludes a comprehensive account of the connection here. An overview of the nature of the connection would be as follows.

The PGC would require that any legal system that sees the status of legal personhood as necessary for the enforcement of rights should ascribe the status to all agents equally; to discriminate on grounds other than bare agency would be to act on an arbitrary characteristic and would not be PGC compliant. This is because, in doing so, the rule denying legal personhood to certain classes of agents (or the absence of an express rule extending it to them) relies on said arbitrary characteristic as a justificatory principle. It is therefore a paradigmatic example of inconsistency, as like cases are not being treated alike. Any rule that claimed to do so would therefore be incapable of providing an absolute reason as to why the restriction should be followed by those operating within the system. If it cannot claim to provide such a reason it cannot be a valid legal rule, meaning that the metaphysical conception of natural personhood²¹ grounded in the bare agency of the subject should be seen a sufficient criterion for the ascription legal personhood. It is therefore a logical necessity that the PGC compliant natural law theory outlined above requires a legal system to recognise that any AI that possesses noumenal agency, or that we believe to do so based on the precautionary principle, must in turn be recognised as a legal person should legal personhood be seen

¹⁴ Many would disagree with the ability of a purely functional account of law to fulfil this purpose. Most recently, see Tamanaha (2017). Such rejections do not engage with the axiom that law must claim to guide human conduct; that other social institutions such as etiquette seek to do the same does not make it less true of law.

¹⁵ Beyleveld and Brownsword (1986), 120. For more on the desirability of such an agreed referent, see Toddington (1993), 58 and Capps (2009), 43.

¹⁶ Capps (2009), 328.

¹⁷ Beyleveld and Brownsword (n 15) 328.

¹⁸ Chen and Burgess (2019), 80, Bartosz Brożek, ‘The Troublesome ‘Person’ in Kurki and Pietrzykowski (2017), 8; Smith (1928), 283.

¹⁹ Kurki (2019) and Naffine (2003).

²⁰ Jowitt (2020, in press) and Jowitt (2019).

²¹ Other conceptions of legal personhood exist; see *inter alia* Naffine, ‘Who are Law’s Persons—from cheshire cats to responsible subjects’ (n 19); Naffine (2009). Attempts to dismiss metaphysics as universally relevant to the recognition of legal personhood do not account for the dialectically necessary argument for the PGC and should be rejected.

as necessary by that system for the recognition and enforcement of legal rights

4 The objection of Bryson, Diamantis, and Grant

If a legal system requires the status of legal personhood for a subject to claim legally enforceable rights, noumenal agency is the overriding factor in ascribing the status to beings other than ourselves. Such agency should be a sufficient, though not necessary, requirement for the recognition of legal personhood in any system where it is necessary that legal personhood be recognised for legal rights to be enforced. Failure to do so would breach the PGC and render any such limitation incapable of possessing the normative force that law necessarily must claim. This is the overriding natural law case for legal personhood for AI that we are required to accept.

Attempts to ground legal personhood in natural personhood through natural law arguments are, however, often rejected. Many, such as Bryson et al.²² have argued that pragmatic concerns about legal personhood for AI mean that any natural law imperative in favour of recognition should be overridden, and their objections will be addressed here in the hope that they can be shown to be misplaced both legally and philosophically. Their concerns can be categorised into three main points. Each will be dismissed in turn, beginning with their first. I will refer to this as the protectionist point:

While not always a zero-sum game, sometimes extending the class of legal persons can come at the expense of the interests of those already within it.²³

The objection is that to extend rights to AI may restrict the ability of human beings to enjoy current legally enforceable rights. This objection is flawed in that it fails to engage with any normative claim for rights for AI, such as that outlined above. To give a clear example as to why this argument fails, we need only realise that it could have been levelled by former slave owners who objected to what was previously considered property being granted legally enforceable rights. These owners' property rights were undoubtedly diminished by granting slaves their freedom, but this was not a good reason to draw the emancipatory project to a close. Natural Law justifications for the recognition of legal personhood such as that defended here are founded in the claim-rights of those to whom personhood is to be extended as opposed to existing classes of legal persons, and it is this point that

Bryson et al. fail to acknowledge. The protectionist argument therefore fails to overcome the claim made by this paper.

Their second category of objections belongs to what appears to be a Realist position:

In many cases, though, the [legal recognition to the rights and obligations of entities that really are people] will not track ... metaphysical and ethical truths. Sometimes legal personhood may be denied to real people to serve odious ends, like perpetuating privileges for some smaller group of People.²⁴

This objection highlights that legal personhood is, in fact, a fiction. It can be granted at the discretion of the legal system in question, and any input as to the desirability of its extension from biology, sociology, ethics or any other external discipline is irrelevant.²⁵ The point is more sophisticated than the first, but remains flawed in that it relies on a notion of personhood grounded in a positivist understanding of legal normativity. Bryson et al. demonstrate a commitment to inclusive legal positivism in their claim that ethical considerations regarding a given rule can only be considered legal if incorporated into the legal system by the legislature, rather than being a necessary condition for the legality of the rule in question.²⁶ Such a commitment is something that is not a satisfactory agreed referent for an inquiry into legal personhood, as it presupposes the conclusion it seeks to establish.²⁷ It is therefore inferior to the Gewirthian account of legal normativity established above, and the necessary conclusion that noumenal agency should be seen as a sufficient, though not necessary, criterion for legal personhood in any system that sees the status as necessary for the recognition of legal rights. This conclusion allows legal systems to retain discretion as to whether they feel the status should be ascribed to non-agents, thus recognising the quasi-fictional element identified by Bryson. Yet it also holds a necessary core of recognition that the authors fail to acknowledge because of their positivist starting point. We are therefore also able to dismiss the second objection as flawed.

Their final objection is functional, in that it seeks to demonstrate that the granting of legal personhood to novel beings is incompatible with the very purpose of a human legal system. By focussing on the function of the phenomenon under consideration, Bryson et al. are attempting to circumvent our dismissal of their second objection and locate an agreed referent from which to construct their claim.²⁸ They hold that the purpose of a human legal system is threefold as follows:

²⁴ Bryson et al. (2017), 278.

²⁵ Bryson et al. (2017), 279.

²⁶ Coleman (2001).

²⁷ Beyleveld and Brownsword (n 15).

²⁸ Toddington(n 15) 58; Capps (n 15) 23, 43.

²² Bryson et al. (2017).

²³ Bryson et al. (2017), 275.

1. to further the material interests of the legal persons it recognizes, and
2. to enforce as legal rights and obligations any sufficiently weighty moral rights and obligations, with the caveat that.
3. should equally weighty moral rights of two types of entity conflict, legal systems should give preference to the moral rights held by human beings.²⁹

Again, this is not a referent upon which both sides would agree; those who would seek to expand personhood to AI provided a threshold criterion was attained would reject the anthropocentricity of point 3) for presupposing the conclusion it seeks to establish. This objection is anticipated by the authors, who attempt to justify the anthropocentricity of their referent by acknowledging it is ‘a kind that allows for deference to the weighty interests of other entities, via the mechanism of human investment in those entities.’³⁰ Yet this justificatory point appears to overlook the normative rights-claims of those beings who may be invested in by human beings; it is grounded in the same protectionist logic as their first objection and can be rejected for the same reason. By extension, there is no reason to accept the anthropocentricity of legal systems presupposed by the referent and the objection fails to adequately counter the normative claim raised by those advocating for legal personhood for AI.

A foundational reason for the failure of this tripartite rebuttal is a reliance on a contested notion of legal personhood. Bryson et al. rely on an extremely narrow definition proposed by Solaiman, that legal personhood can only be granted to a being ‘that it is able to know and execute its rights as a legal agent, and that it is subject to legal sanctions ordinarily applied to humans.’³¹ The authors believe that AI would be incapable fulfilling the second criterion which, if accepted, could justify their anthropocentric conception of law. The criterion should be unequivocally rejected, however. Firstly, in introducing the arbitrary criterion of being subject to a legal sanction, it does not overcome the central claim of this paper—that legal personhood is not a mere artifice that can be ascribed at the discretion of a legal system, but is inseparable from natural personhood possessed through bare agency. Second, it is being blind to the fact that things that are incapable of being subject to sanction are already granted the status of legal person. Rivers have already been granted this status in New Zealand³²

and India,³³ and it would be conceptually impossible for these entities to be subject to sanctions ordinarily applied to humans. Similarly, the Criminal Law of many jurisdictions holds children to be incapable of being subject to legal sanctions due to their age, but this does not remove their status as legal persons.³⁴ These examples show that Solaiman is focussing excessively on the notion of active over passive legal personhood, prioritising an ability to interact with other legal persons over the ability to be recognised as being in possession of legal rights even if these rights cannot be claimed by the legal person themselves.³⁵ As passive legal personhood detached from an active element is recognised as being sufficient for the recognition of legal personhood proper by several contemporary legal systems, the criterion held to be necessary by Solaiman is anything but and should be rejected. A being need not be capable of being subject to sanction for their legal personhood to be recognised. In the absence of a satisfactory counterargument, bare agency should be recognised as a sufficient though not necessary criterion for the recognition of legal personhood in systems that require the status for rights to be legally enforced.

5 Compatibility of current legislative proposals

As with all shifts in the scope of legal recognition of rights interests, the argument that has been presented is one that will take time to be accepted. A long struggle by human advocates for rights for AI may also be anticipated, with all the difficulties associated with extending rights to things traditionally seen as property.³⁶ Yet AI is already impacting the law, even before it can be seen as a legitimate rights holder—and the law is already reforming to account for this. In insurance, AI is already outpacing legal rules and the law frequently has to catch up with technological developments.³⁷ In Contract Law, AI can already enter into contracts but debates rage as to the location of the intentionality behind the agreement,³⁸ and it has been argued that AI already possess a level of cognition to have rules of Tortious liability applied to their actions.³⁹

²⁹ Bryson (n 22) 283.

³⁰ Coleman (2001), 283.

³¹ Coleman (2001), 274 and Solaiman (2017).

³² Te Awa Tupua (Whanganui River Claims Settlement) Act (2017).

³³ *Mohd. Salim v State of Uttarakhand and others* (Writ Petition (PIL) No.126 of 2014).

³⁴ Children and Young Persons Act 1933 s 50, as amended by Children and Young Persons Act 1963 s 16 (1).

³⁵ For more on the notion of active v passive legal personhood, see Kurki (2019).

³⁶ Zenor (2018), 126.

³⁷ Cowgill (2019).

³⁸ Dahiyat (2007).

³⁹ Sartor (2009).

Given the rapidity with which such technologies are developing, it is essential that the law move now to address how AI should be accounted for in law. Experience from the fantasy sports and tobacco industries suggests that it is desirable that this take place now, as market pressure generated by product users means that developing industries are easier to regulate than ones that are already established.⁴⁰ This is a more pressing concern for the granting of legally enforceable rights than for traditional forms of regulation, given that failure to adequately protect the legitimate claim-rights of AI would be a breach of the moral imperative of the PGC. With this in mind, the final section of this article will examine current legislative proposals in the UK and European Union in turn with the aim of identifying the extent to which they successfully future-proof themselves by adequately addressing the case for legal personhood for AI constructed above.

5.1 The United Kingdom

In a recent House of Commons Science and Technology Committee Report, legislative attitude can be identified by the focus of the report on how developments in AI may affect extant social norms. Language such as ‘Machines versus humans’⁴¹ demonstrate that the Committee is committed to an anthropocentric view of legal personhood, and in doing so have made the same conceptual error as Solaiman and Bryson et al. They do not consider the possibility that AI may, in future, meet the threshold criterion of noumenal agency and become bearers of legitimate claims to legal protection of those rights against human interference. In this regard, they are acting analogously to the slave-owner who cannot conceptualise the possibility that his possessions may be legitimate rights holders. This inability to countenance the possibility that AI may be legitimate rights-holders in the future is myopic, and a failure to adequately future-proof any legislation that might be passed with regards to the regulation of AI.

A recent report of the House of Lords Select Committee on Artificial Intelligence is guilty of the same unconscious anthropocentricity. The word ‘rights’ is used only sixteen times in a document of almost two-hundred pages, and these uses are all focussed on the impact AI may have on the existing rights of human persons.⁴² The issue of AI as legal persons was raised by Dr Sarah Morley and Dr David

Lawrence of Newcastle Law School, and although the committee acknowledged that this is an issue that could be legislated upon, the desirability of this step was not addressed.⁴³ Like their Commons counterparts, the Lords Committee has failed to provide suggestions that future-proof legislation in a way that can account for the possibility that AI may come to possess the noumenal agency required for PGC compliant legal personhood. The jurisdictions of the UK may fall foul of this short-sightedness by creating a situation in which legitimate rights-claims made by AI may be overridden by human interests without due consideration as to which interests are of a higher normative significance.

It is notable to contrast this legislative blindness with the openness to the possibility of such personhood from the judicial branch; the subject was addressed directly by Lord Hodge at a recent lecture given at the University of Edinburgh. Acknowledging that much legal time has been spent considering the ‘attribution of responsibility for the acts and omissions of robots,’⁴⁴ he continued to the issue of personhood for computers. Given that English Law has recognised the personality of ruined temples in foreign jurisdictions,⁴⁵ he claimed [‘T]here is no reason in principle why the law cannot create such personality’ for Artificial Intelligence.⁴⁶ Indeed, he offered detailed proposals:

It would be possible for the machine as a separate legal person to own intellectual property and in turn to be owned by a financial institution. That institution’s licence or the general regulatory law could impose on the firm responsibility for any malfunction, if, for example, it had been involved in the design of the algorithm. The law could confer separate legal personality on the machine by registration and require it or its owner to have compulsory insurance to cover its liability to third parties in delict (tort) or restitution. And as a registered person the machine could own the intellectual property which it created.⁴⁷

Lord Hodge’s proposals are therefore similar to the guardianship model already used by several North American municipalities as a means to recognise that animals kept as pets are legitimate rights-holders beyond mere property,⁴⁸ and adopted as a means of protecting the rights of rivers previously mentioned as possessing personhood in both New

⁴⁰ Weaver (2018).

⁴¹ UK Parliament, Science and Technology Committee, Robotics and Artificial Intelligence (HC 2016–17, 145) 9.

⁴² UK Parliament, House of Lords Select Committee on Artificial Intelligence, AI in the UK: ready, willing and able? (HL 2017–19, 100).

⁴³ Weaver (2018), 99, ¶314.

⁴⁴ Hodge (2019).

⁴⁵ Bumper Development Corporation v Commissioner of Police of the Metropolis (1991).

⁴⁶ Hodge (n 44), 17.

⁴⁷ Bumper Development Corporation v Commissioner of Police of the Metropolis (1991).

⁴⁸ Sullivan and Vietzke (2008), 44–45.

Zealand⁴⁹ and India.⁵⁰ Such a development, were it to be adopted in the UK, would be a desirable recognition of the legal personhood required to protect the rights of AI. Should the Courts see a reason to drive the change, the Common Law may serve as a vehicle for the recognition of legal personhood for AI on a case by case basis where the legislature is reluctant to act.⁵¹ Such judicial activism may eventually be a catalyst for legislative action to address the issue comprehensively rather than on a piecemeal basis.

5.2 The European Union

EU legislative proposals expressly recognise that the law is playing catch-up,⁵² but fall foul of the same unjustified anthropocentricity that is seen in the UK legislature. The Commission recognises that ‘Europe is well placed to establish a distinctive form of AI that is ethically robust and protects the rights of individuals, firms, and society at large.’⁵³—but AI is itself absent in the latter list of those whose rights ought to be protected. The Commission proposals are therefore blind to the possibility that AI may, now or in the future, be legitimate rights holders.⁵⁴ Similarly, Commission research identifies a focus is needed on the autonomy, identity and dignity of AI users and ethical issues that arise from this,⁵⁵ whilst remaining silent on the point of whether there should be equal consideration for the same qualities if present in AI.

The lack of consideration for the legal rights claims of AI is also present in the Commission’s High Level Expert Group on AI, but instead of being a blind spot on the part of the proposers, such rights are explicitly rejected. It expressly states that AI systems need to be ‘human-centric, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom.’⁵⁶ This is problematic in light of the argument presented in the first part of this paper, as the primacy of the human being is merely assumed as being of central importance without any attempt at providing a normative foundation for this position:

The common foundation that unites [existing human rights] can be understood as rooted in respect for human dignity – thereby reflecting what we describe as a “human-centric approach” in which the human being enjoys a unique and inalienable moral status of primacy in the civil, political, economic and social fields.⁵⁷

Whilst no justification of the ‘unique status’ claimed is given here, the Expert Group’s guidelines on ethics and AI contain the following justification for maintaining this primacy:

The human-centric approach to AI strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights, including those set out in the Treaties of the European Union and Charter of Fundamental Rights of the European Union, all of which are united by reference to a common foundation rooted in respect for human dignity, in which the human being enjoy a unique and inalienable moral status. This also entails consideration of the natural environment and of other living beings that are part of the human ecosystem, as well as a sustainable approach enabling the flourishing of future generations to come.⁵⁸

This glossary definition for human centric AI is problematic for two reasons. Firstly, it is unclear how the group can reconcile its human centric approach with the consideration of other ‘living beings’ in the final section of the definition. This clause is clearly inserted to demonstrate that such beings have been considered, but the lack of normative engagement with their status and how any conflict between human beings and other living beings would be resolved shows that this is mere puffery. Secondly, the fact this consideration is limited to living beings shows the authors are unwilling to address the issue of rights for AI. On a normative level, this biocentrism is as problematic as the anthropocentrism of other EU bodies and efforts should be made to address the omission.

EU bodies other than the Commission are similarly anthropocentric. A recent EU Parliament Resolution focussed entirely on harm that AI can do, not the status of any interaction between AI and human beings,⁵⁹ and the EU Declaration of Cooperation on AI explicitly states, with no attempt at normative justification, that its aim is to:

⁴⁹ Te Awa Tupua (Whanganui River Claims Settlement) Act 2017 ss 18–20 (2) (New Zealand).

⁵⁰ *Mohd. Salim* (n 33), ¶19.

⁵¹ Atiyah and Summers (1987), 134, Llewellyn (1960), 90–91 and Kelch 1998), 546.

⁵² European Commission (2019a).

⁵³ European Commission (2019a), 10.

⁵⁴ European Commission (2019a), 8.

⁵⁵ European Commission (2019a), s.6.2.1.

⁵⁶ European Commission (2019b).

⁵⁷ European Commission (2019b), 10.

⁵⁸ European Commission (2019b), 37.

⁵⁹ European Parliament (2017).

Ensure that humans remain at the centre of the development, deployment and decision-making of AI, prevent the harmful creation and use of AI applications, and advance public understanding of AI.⁶⁰

Were the subject of this sentiment a child or other vulnerable minority group, then it would rightly be considered indefensible, as such a position does not engage with the normative element of rights protection. It should be seen as equally indefensible to not consider the normative issue raised by AI that make legitimate claim-rights under the PGC. The EU can therefore be criticised for the same reasons that Lawrence and Brazier have criticised other rights documents; it rests on a presumed notion that personhood is inherently anthropocentric, and the status of ‘rights-bearer’ is fixed and immovable as it was settled before current technology existed.⁶¹ The purpose of this article has been to demonstrate that his assumption is false. AI that are subject to the requirements of the PGC should be seen as legitimate rights-holders, and it is a necessary condition of a functioning legal system that this be recognised by any system that requires recognition of legal personhood for the enforcement of rights. Any serious attempt at law reform to account for such beings must therefore engage with this issue seriously, rather than avoid it for political expediency.

6 Conclusion

This paper does not seek to argue that extant AI meets the threshold criterion of agency required for the recognition of legal personhood. Rather, it asks the question looking to a future where the threshold criterion may be reached. Having established the moral case for legal personhood grounded in a strong conception of natural law, we can see that objections grounded in self-preservation or pragmatism do not overcome the normative claim contained in the moral case that has been developed. As such, any legal system should seek to future proof itself by accounting for the possibility that AI may reach the threshold criterion of agency that has been identified as a sufficient, though not necessary, criterion of legal personhood in any legal system where legal personhood is deemed necessary for the ascription of legal rights. This is a future that is not inconceivable, and it is better that legal systems prepare themselves for it now rather than find themselves in breach of the moral imperative that has been developed. To fail to do so would be to knowingly revisit the mistakes of the past, in which the legal personhood of several classes of people were knowingly removed

based on arbitrary criteria such as race and gender.⁶² Only the recognition of the metaphysical core of legal personhood can help us to avoid revisiting these past injustices against new categories of beings.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Atiyah PS, Summers R (1987) Form and substance in Anglo-American Law. Clarendon Press, Oxford
- Beyleveld D (1991) The dialectical necessity of morality; an analysis and defense of Alan Gewirth’s argument to the principle of generic consistency. University of Chicago Press, Chicago
- Beyleveld D (2012) The principle of generic consistency as the supreme principle of human rights. *Hum Rights Rev* 13:1
- Beyleveld D, Brownsword R (1986) Law as a moral judgment. Sweet and Maxwell, London
- Bryson J et al (2017) Of, for, and by the people: the legal lacuna of synthetic persons. *Artif Intell Law* 25:273
- Bumper Development Corporation v Commissioner of Police of the Metropolis (1991) 1 WLR 1362
- Capps P (2009) Human dignity and the foundations of international law. Hart, Oxford
- Chen J, Burgess P (2019) The boundaries of legal personhood: how spontaneous intelligence can problematise differences between humans, artificial intelligence, companies and animals. *Artif Intell Law* 27(1):73
- Children and Young Persons Act (1933)
- Children and Young Persons Act (1963)
- Coleman J (2001) The practice of principle: in defence of a pragmatist approach to legal theory. Oxford University Press, Oxford
- Cowgill A (2019) Artificial intelligence: a grey area for insurance coverage. *J Robot Artif Intell Law* 1(1):35
- Dahiyat EAR (2007) Intelligent agents and contracts: is a conceptual rethink imperative? *Artif Intell Law* 15:375
- Dredd Scott v Sandford 60 US 393 (1856)
- European Commission (2019) European Commission artificial intelligence: a European perspective (Joint Research Centre) JRC 113826 <http://publications.jrc.ec.europa.eu/repository/handle/JRC113826>. Accessed 10 Apr 2019
- European Commission (2019) European Commission—independent high level expert group on AI ‘Ethics Guidelines for Trustworthy AI’ https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477. Accessed 10th Apr 2019

⁶⁰ European Union (2018).

⁶¹ Lawrence and Brazier (2018), 320.

⁶² The lack of recognition of slaves as legal persons would be a paradigm example. See *Dredd Scott v Sandford* 60 US 393 (1856).

- European Parliament (2017) European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))
- European Union (2018) Declaration of cooperation on artificial intelligence. <https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence>. Accessed 10 Apr 2019
- Gewirth A (1978) Reason and morality. University of Chicago Press, Chicago
- Gewirth A (1982) Human rights: essays on justification and applications. University of Chicago Press, Chicago
- Gewirth A (1996) The community of rights. University of Chicago Press, Chicago
- Hodge L (2019) The potential and perils of financial technology: can the law adapt to cope? (The First Edinburgh FinTech Law Lecture, University of Edinburgh, 14 March 2019). <https://www.supremecourt.uk/docs/speech-190314.pdf>. Accessed 19 Mar 2019
- Jowitt J (2019) Legal rights for animals: aspiration or logical necessity?. IVR World Congress, Lucerne
- Jowitt J (2020) The desirability of legal rights for novel beings. Cambridge Quarterly of Healthcare Ethics (**In press**)
- Kelch TG (1998) Toward a non-property status for animals. N Y Univ Environ Law J 6(3):531
- Kurki V (2019) A theory of legal personhood. Oxford University Press, Oxford
- Kurki V, Pietrzykowski T (eds) (2017) Legal personhood: animals, artificial intelligence and the unborn. Springer, Cham
- Lawrence D, Brazier M (2018) Legally human? ‘Novel Beings’ and English Law. Med Law Rev 26(2):309
- Llewellyn K (1960) The common law tradition: deciding appeals. Little, Brown and Co, New York
- Mohd. Salim v State of Uttarakhand and others (Writ Petition (PIL) No. 126 of 2014) para 19 <https://drive.google.com/file/d/0BzXilfcx7yuM3VRWTZDeEtmSGc/view>. Accessed 26 Mar 2019
- Naffine N (2003) Who are Law’s persons—from cheshire cats to responsible subjects. Mod Law Rev 66:346
- Naffine N (2009) Law’s meaning of life: philosophy, religion, darwin and the legal person. Hart, Oxford
- Nagel T (1974) What is it like to be a bat? Philos Rev 83(4):435
- UK Parliament (2016–17) Science and Technology Committee, robotics and artificial intelligence (HC 2016-17, 145) <https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2015/robotics-and-artificial-intelligence-inquiry-15-16/>. Accessed 17 Mar 2019
- UK Parliament (2017–19) House of Lords Select Committee on Artificial Intelligence, *AI in the UK: ready, willing and able?* (HL 2017-19, 100) <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>. Accessed 17 Mar 2019
- Regis E Jr (ed) (1984) Gewirth’s ethical rationalism. University of Chicago Press, Chicago
- Sartor G (2009) Cognitive automata and the law: electronic contracting and the intentionality of software agents. Artif Intell Law 17:253
- Sebo J (2018) The moral problem of other minds Harv Rev Philos 25:51
- Smith B (1928) Legal personality. Yale Law J 37(3):283
- Solaiman SM (2017) Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy. Artif Intell Law 25(2):155
- Solum L (1992) Legal personhood for artificial intelligences. N C Law Rev 70(4):1231
- Sullivan D, Vietzke H (2008) An animal is not an iPod. J Anim Law 4:41
- Tamanaha B (2017) A realistic theory of law. Cambridge University Press, Cambridge
- Te Awa Tupua (Whanganui River Claims Settlement) Act 2017 (New Zealand)
- Toddington S (1993) Rationality. Social Action and Moral Judgment. Edinburgh University Press, Edinburgh
- Weaver J (2018) Everything is not terminator: the importance of regulating AI as soon as possible. J Robot Artif Intell Law 1(2):131
- Zenor J (2018) Endowed by their creator with certain rights: the future rise of civil rights for artificial intelligence? Savannah Law Rev 5(1):115

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.